МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ФИЗИЧЕСКИХ И ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ И ТЕХНИЧЕСКИХ СИСТЕМ

УДК 519.14

Г. С. АБРАМОВ
Херсонский Национальный технический университет
И. М. АБРАМОВ
Херсонская специализированная школа №30, МАН Украины

ГАУССОВСКАЯ АППРОКСИМАЦИЯ ДЛЯ РАСПРЕДЕЛЕНИЯ ЧИСЕЛ В СТРОКАХ ТРЕУГОЛЬНИКА ПАСКАЛЯ

Получены значения параметров нормального распределения, аппроксимирующего распределение чисел в n-ой строке треугольника Паскаля. Вычислены значения нормированных моментов чётных порядков (с 4-го по 8-ой) и показано их асимптотическое приближение к значениям, соответствующим нормальному распределению. Получены высокоточные приближения для центральных элементов чётных строк треугольника Паскаля, что даёт возможность вычислять биномиальные, а также триномиальные (и, в общем случае, мультиномиальные) коэффициенты.

Ключевые слова: треугольник Паскаля, биномиальные коэффициенты, нормальное (гауссовское) распределение, триномиальные и мультиномиальные коэффициенты.

Г. С. АБРАМОВ
Херсонський Національний технічний університет
І. М. АБРАМОВ
Херсонська спеціалізована школа №30, МАН України

ГАУСІВСЬКА АПРОКСИМАЦІЯ ДЛЯ РОЗПОДІЛУ ЧИСЕЛ У РЯДКАХ ТРИКУТНИКА ПАСКАЛЯ

Отримано значення параметрів нормального розподілу, що апроксимує розподіл чисел в n-ому рядку трикутника Паскаля. Обчислені значення нормованих моментів парних порядків (з 4-го по 8-ий) і показано їх асимптотичне наближення до значень, відповідним нормальному розподілу. Отримано високоточні наближення для центральних елементів парних рядків трикутника Паскаля, що дає можливість обчислювати біноміальні, а також триноміальні (і, в загальному випадку, мультиноміальний) коефіцієнти.

Ключові слова: трикутник Паскаля, біноміальні коефіцієнти, нормальний (гаусівський) розподіл, триноміальні та мультиноміальні коефіцієнти.

G. S. ABRAMOV
Kherson National Technical University
I. M. ABRAMOV
Kherson specialized school №30, MAS Ukraine

GAUSS APPROXIMATION FOR NUMBER DISTRIBUTION IN ROWS OF A PASCAL’S TRIANGLE

The parameters of normal distribution which approximate the number distribution in the n-th row of the triangle is determined. The normalized values of even power moments (4 to 8) are found, and their asymptotic approximation to the values which are in accordance to the normal distribution is shown. High-precision approximations for central elements of even rows of a Pascal’s triangle are obtained, which gives the opportunity to calculate the binomial and trinomial coefficients (as well as multinomial ones in general cases).

Keywords: Pascal’s triangle, binomial coefficients, normal (Gauss) distribution, trinomial and multinomial coefficients.

Постановка проблемы

Треугольник Паскаля составляют так называемые биномиальные коэффициенты. Биномиальные и триномиальные коэффициенты часто встречаются в различных технических приложениях. Так, во многих задачах физической и химической кинетики, а также в задачах теории информации возникает необходимость вычислять различные конфигурации на множествах, в частности числа сочетаний C_n^m, то есть биномиальные коэффициенты. Чаще всего в этих случаях для вычисления факториалов используют приближённую формулу Стирлинга, которая даёт достаточно грубое приближение. Поэтому задача отыскания более точного приближения является актуальной.

Известно также, что асимптотикой биномиального распределения является нормальное (гауссовское) распределение. Однако оно реализуется лишь в пределе. Конкретные же сведения о «кинетике» этого приближения; о том, каким образом изменяются числовые характеристики аппроксимирующего нормального распределения при увеличении номера строки, отсутствуют.

Анализ публикаций

Треугольнику Паскаля посвящена довольно обширная литература [2-6], в которой выделяются две фундаментальные монографии [2, 3], где биномиальным коэффициентам посвящены отдельные главы. В книге Р. Грэхема, Д. Кнута и О. Паташника «Конкретная математика. Основание информатики» [2] в главе «Биномиальные коэффициенты» рассматриваются доказательства многочисленных (часто удивительных) тождеств и оригинальные способы вычисления различных частных сумм. В монографии В. Феллера «Введения в теорию вероятностей и её приложения» [3] рассматривается нормальное приближение для биномиального распределения, однако погрешность результатов (1-2%) нельзя признать удовлетворительной.

Формулировка целей исследования

Целями данной статьи, являющейся продолжением исследований, начатых в работе [1], было рассмотрение асимптотики приближения распределения чисел в строках треугольника Паскаля к нормальному; вычисление нормированных моментов чётных порядков и изучение асимптотики их изменения при увеличении номера строки; рассмотрение отклонений плотности паскалевского распределения от гауссовского в окрестности центральных элементов; получение высокоточных приближений для центральных элементов строк треугольника Паскаля и их использование для вычисления триномиальных (мультиномиальных) коэффициентов.

Основная часть

В предыдущей работе авторов [1], целью которой было установление связи распределения чисел в строках треугольника Паскаля с нормальным (гауссовским) распределением, были решены следующие задачи: определены числовые характеристики нормального приближения для распределения чисел в строках треугольника Паскаля; выявлены зависимости этих параметров от номера строки; выполнено сравнение распределения чисел в строках треугольника Паскаля с нормальным приближением и рассмотрена асимптотика приближения паскалевского распределения к нормальному при увеличении номера строки.

Математическое ожидание ($\bar{x}$) и дисперсию (S^2) распределения чисел в строках треугольника Паскаля вычисляли следующим образом:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i; \qquad S^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 n_i - \bar{x}^2,$$

где x_i – номер позиции (элемента) в строке треугольника Паскаля;
n_i – значение элемента на i-ой позиции в строке треугольника Паскаля;
$k = n+1$– количество элементов в n-ой строке треугольника Паскаля;
N – сумма всех элементов в данной строке треугольника Паскаля: $N=2^n$.

Математическое ожидание $\bar{x}$, дисперсия S^2 и среднеквадратичное отклонение S от номера строки треугольника Паскаля зависит следующим образом:

$$\bar{x} = \frac{n+2}{2}; \quad S^2 = \frac{n}{4}; \quad S = \frac{\sqrt{n}}{2}.$$

Для соответствующих значений матожидания и среднеквадратичного отклонения построены гауссовские распределения и выполнено сравнение полученных гауссовских кривых с фактическим частотным распределением в соответствующих строка треугольника Паскаля. Для этого были вычислены относительные частоты распределения чисел в строках треугольника Паскаля (значение данного элемента строки делилось на сумму всех чисел в строке – 2^n).

В [1] приведены соответствующие эмпирические распределения (основанные на вычисленных частотах) в сравнении с теоретическими распределениями (гауссовские кривые). Сравнения показало, что с увеличением номера строки распределение чисел в строках треугольника Паскаля приближается к нормальному распределению. Это даёт основание считать, что нормальное приближение является хорошей аппроксимацией распределения в n-той строке треугольника Паскаля (очевидно, тем лучшей, чем больше номер строки n):

$$x_n \in N\left(\frac{n+2}{2}; \frac{\sqrt{n}}{2}\right).$$

Известно, что искажения нормального распределения связаны с моментами третьего и четвёртого порядков (соответствующие коэффициенты называются асимметрией и эксцессом). Асимметрия, описывающая искажения гауссовской кривой вдоль оси абсцисс, в нашем случае отсутствует, в силу симметричности чисел в строках треугольника Паскаля. (Отсюда следует, что все моменты нечётных порядков для строк треугольника Паскаля равны нулю). Следовательно, для описания наблюдаемых различий паскалевского распределения от гауссовского приближения, необходимо вычислять значения эксцесса (связанного с моментом четвёртого порядка), и, вообще говоря, моменты чётных порядков (4-го, 6-го, 8-го и т. д.). Очевидно, что вычисляемые значения эксцесса должны быть отрицательными (т. к. «уплощение» гауссовской кривой характеризуется отрицательным значением эксцесса), при этом, с возрастанием номера строки значение эксцесса должно асимптотически приближаться к нулю.

Эксцесс (E_x) вычисляли с помощью центрального момента 4-го порядка - μ^4:

$$\mu^4 = \frac{\sum_{i=1}^{n+1}(x_i - \bar{x})^4 n_i}{\sum_{i=1}^{n+1} n_i} = \frac{\sum_{i=1}^{n+1}(x_i - \bar{x})^4 n_i}{2^n}; \qquad E_x = \frac{\mu^4}{\sigma^4} - 3$$

В работе [1] показано, что значения эксцесса отрицательны и с увеличением номера строки асимптотически приближаются к нулю, что свидетельствует о приближении распределения в строках с большим номером к гауссовскому.

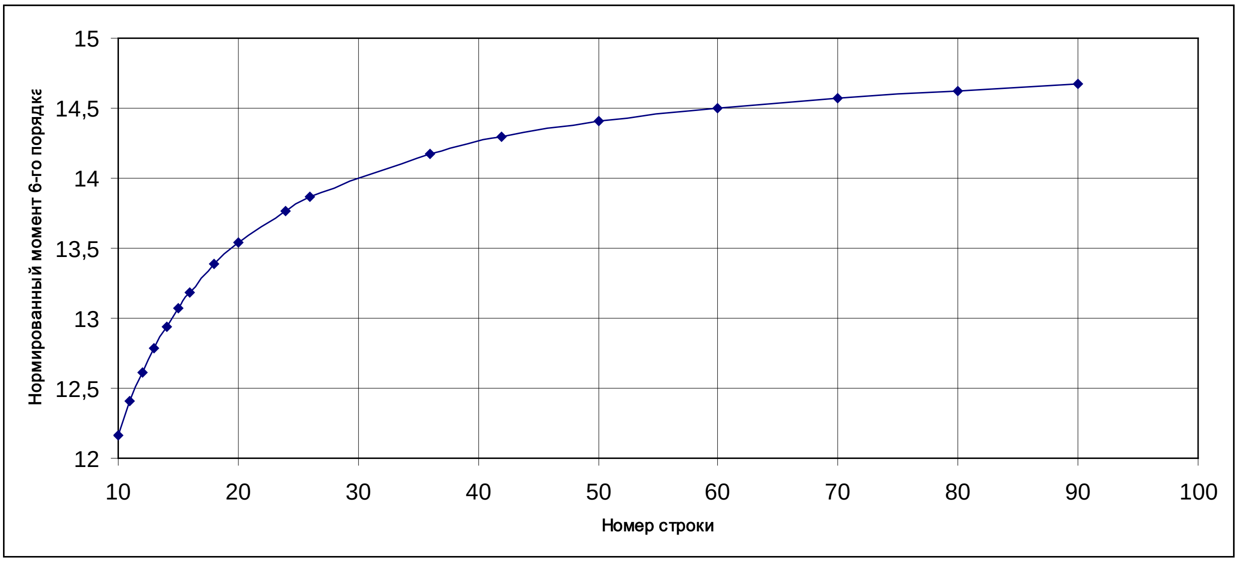

Рис. 1. График зависимости нормированного момента 6-го порядка от номера строки

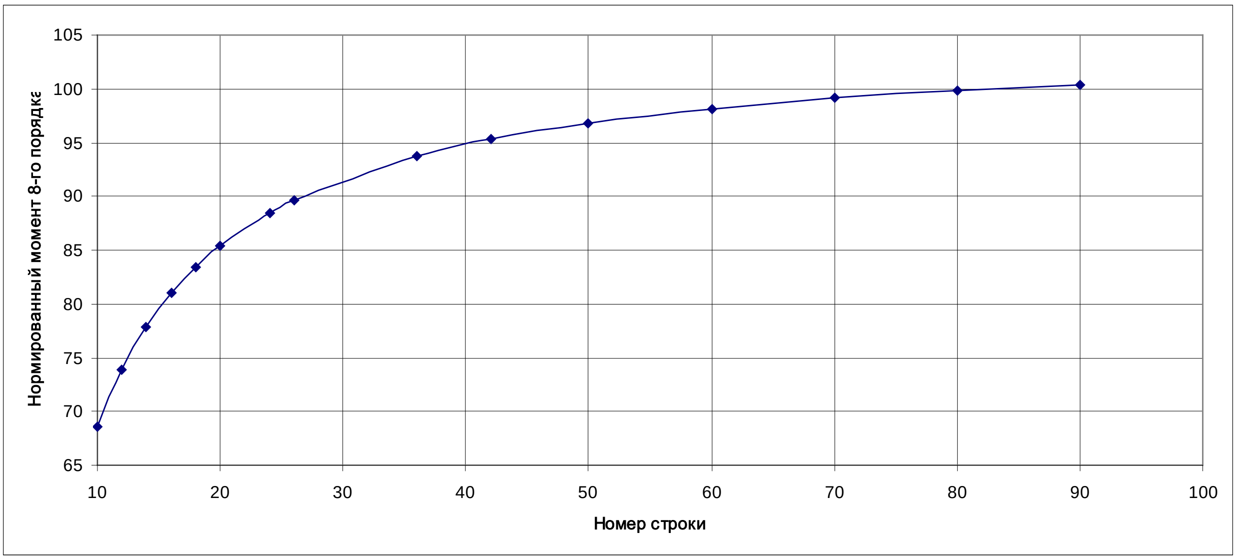

Рис. 2. График зависимости нормированного момента 8-го порядка от номера строки

В настоящей работе рассматривались моменты 6-го и 8-го порядков распределения в строках треугольника Паскаля. На рис. 1 представлена зависимость нормированного момента 6-го порядка $\frac{\mu^6}{\sigma^6}$ от номера строки, а на рис. 2 представлена зависимость нормированного момента 8-го порядка $\frac{\mu^8}{\sigma^8}$ от номера строки. Для нормального распределения значение нормированного момента 6-го порядка равно 15, а 8-го порядка равно 105. Видно, что при увеличении номера строки значения нормированного момента 6-го порядка асимптотически приближаются к 15, а 8-го порядка к 105, со стороны меньших значений, то есть распределение в строках треугольника Паскаля нормализуется и с точки зрения моментов 6-го и 8-го порядков (по-видимому, это будет справедливо и для последующих чётных моментов).

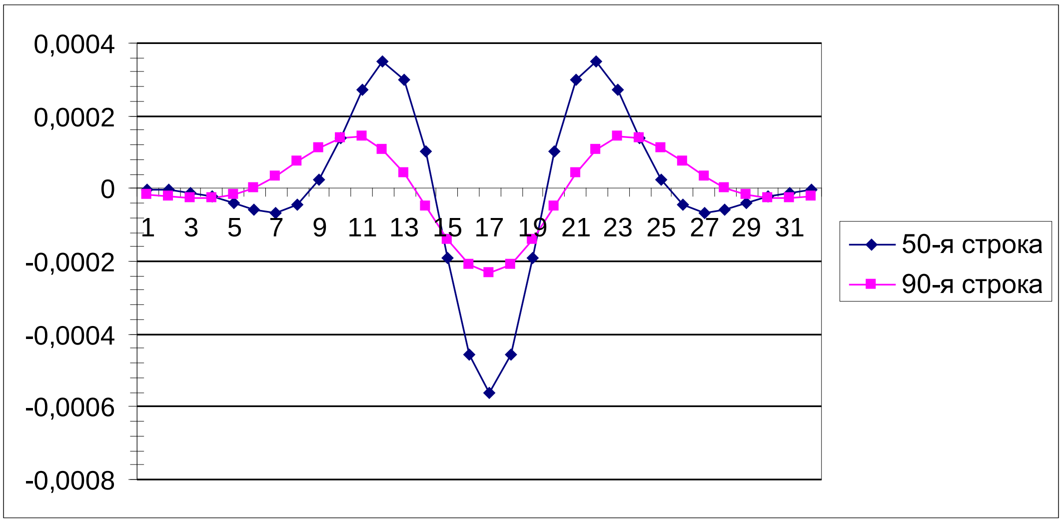

Рис. 3 Графики отклонений значений плотности распределений в строках треугольника Паскаля от нормального распределения в окрестности центральных элементов для 50-ой и 90-ой строк

На рис. 3 представлены графики отклонений значений плотности распределений в строках треугольника Паскаля от нормального распределения в окрестности центральных элементов для 50-ой и 90-ой строк.

Видно, что если для центральных элементов строк эта разность отрицательна (паскалевское распределение «приплюснуто» относительно гауссовского), то в ближайшей окрестности наблюдается обратная ситуация – плотность паскалевского распределения больше, чем плотность гауссовского распределения. Дальше, с удалением от центральных элементов снова наблюдается отрицательная разность, однако этот «волнообразный» процесс быстро затухает и разность асимптотически приближается к нулю.

Видно также, что для строки с более высоким номером амплитуда колебаний меньше по сравнению с распределением в строке с меньшим номером, то есть с возрастанием номера строки идёт сглаживание (затухание) колебаний плотности паскалевского распределения относительно гауссовского и с точки зрения их амплитуды, что ещё раз показывает – при увеличении номера строки паскалевское распределение приближается к гауссовскому.

Таким образом, гауссовское распределение является весьма хорошей аппроксимацией для паскалевского распределения в строках треугольника. Плотность распределения в n-ой строке треугольника имеет следующий вид:

$$f_n(x) \approx \frac{2}{\sqrt{2\pi n}} \exp\left(-\frac{\left(x - \frac{n+2}{2} \right)^2}{\frac{2n}{4}} \right) = \sqrt{\frac{2}{\pi n}} \exp\left(-\frac{2\left(x - \frac{n+2}{2} \right)^2}{n} \right). \quad (1)$$

В работе [1] с использованием формулы (1) проведены приближённые вычисления элементов строк треугольника Паскаля. Число сочетаний C_n^m в нашем случае является $(m+1)$-м элементом n-ой строки треугольника (т. к. отсчёт величины x мы начали не с нуля а с единицы). Тогда число сочетаний можно вычислить по следующей приближённой формуле:

$$C_n^m \approx 2^n \sqrt{\frac{2}{\pi n}} \exp\left(-\frac{(2m-n)^2}{2n} \right). \quad (2)$$

Проще всего с помощью этой формулы вычисляется значение центральных элементов чётных строк треугольника, когда $m=n/2$:

$$C_n^{m=\frac{n}{2}} \approx \frac{2^n}{\sqrt{\pi n}} \text{ или } C_{2n}^n \approx \frac{4^n}{\sqrt{\pi n}} \qquad (3)$$

Можно показать, что приближение (3) соответствует вычислению числа сочетаний, если факториалы вычислять по формуле Стирлинга:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \ . \qquad (4)$$

В работе [1] проведено сравнение точных значений центральных элементов чётных строк треугольника и их приближённых значений, вычисленных с помощью гауссовской аппроксимации (табл. 1):

Таблица 1

Сравнение точных значений и гауссовской оценки центральных элементов чётных строк

Номер строки n	Точное значение $C_n^{m=\frac{n}{2}} = \frac{n!}{(m!)^2}$	Приближённое значение $C_n^{m=\frac{n}{2}} \approx \frac{2^n}{\sqrt{\pi n}}$	Относительная ошибка ε, %
30	$1{,}551175 \cdot 10^{8}$	$1{,}564153 \cdot 10^{8}$	0,84
50	$1{,}264106 \cdot 10^{14}$	$1{,}270442 \cdot 10^{14}$	0,50
60	$1{,}182646 \cdot 10^{17}$	$1{,}187583 \cdot 10^{17}$	0,42
100	$1{,}008913 \cdot 10^{29}$	$1{,}011439 \cdot 10^{29}$	0,25
200	$9{,}054851 \cdot 10^{58}$	$9{,}066177 \cdot 10^{58}$	0,125
400	$1{,}029525 \cdot 10^{119}$	$1{,}030169 \cdot 10^{119}$	0,0625
800	$1{,}88042442 \cdot 10^{239}$	$1{,}88101214 \cdot 10^{239}$	0,03125

Из данных табл. 1 следует, что относительная ошибка (разность между приближённым и точным значением, отнесенная к точному значению) убывает с возрастанием номера строки по обратно пропорциональной зависимости, что позволяет записать выражение для относительной ошибки:

$$\frac{\frac{2^n}{\sqrt{\pi n}} - C_n^{m=n/2}}{C_n^{m=n/2}} = \frac{0{,}25}{n}, \qquad (5)$$

откуда следует более точное приближение:

$$C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi n}} \cdot \frac{n}{n+0{,}25} \ . \qquad (6)$$

Сравним точное значение с приближением (6) (табл. 2):

Таблица 2

Сравнение точных значений и приближения (6) для центральных элементов чётных строк треугольника Паскаля

Номер строки n	Точное значение $C_n^{m=\frac{n}{2}} = \frac{n!}{(m!)^2}$	Приближённое значение (6) $C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi n}} \cdot \frac{n}{n+0{,}25}$	Относительная ошибка ε_1, %
30	$1{,}551175 \cdot 10^{8}$	$1{,}551226 \cdot 10^{8}$	0,0033
50	$1{,}264106 \cdot 10^{14}$	$1{,}264121 \cdot 10^{14}$	0,0012
60	$1{,}182646 \cdot 10^{17}$	$1{,}182655 \cdot 10^{17}$	0,0008
100	$1{,}0089134 \cdot 10^{29}$	$1{,}0089166 \cdot 10^{29}$	0,0003
200	$9{,}0548515 \cdot 10^{58}$	$9{,}0548578 \cdot 10^{58}$	0,00007
400	$1{,}029525 \cdot 10^{119}$	$1{,}029525 \cdot 10^{119}$	$\varepsilon_1 = \frac{3}{n^2}\%$
800	$1{,}88042442 \cdot 10^{239}$	$1{,}88042442 \cdot 10^{239}$	

Видно, что относительная ошибка этого приближения ε_1 обратно пропорциональна квадрату номера строки, что даёт возможность сформулировать ещё более точное приближение:

$$C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi n}} \cdot \frac{n}{n+0{,}25} \cdot \frac{n^2}{n^2+0{,}03} \qquad (7)$$

или

$$C_{2n}^n \approx \frac{4^n}{\sqrt{\pi n}} \cdot \frac{n}{n+0{,}125} \cdot \frac{n^2}{n^2+0{,}0075} \ . \qquad (8)$$

Точность этого приближения настолько высока, что даже для небольших n она даёт очень хорошую оценку. Так, число сочетаний C_{20}^{10}=184756 , а формулы (7-8) дают результат 184755,5; т. е. относительная ошибка составляет 0,00027%.

Формула (8) может быть записана в виде:

$$C_{2n}^{n} \approx \frac{4^n}{\sqrt{\pi n}} \cdot \left(\frac{1}{1+\frac{0,125}{n}+\frac{0,0075}{n^2}+\frac{0,0009375}{n^3}} \right) \qquad (9)$$

Имея достаточно точные значения центрального элемента чётной строки треугольника Паскаля можно легко вычислить и остальные элементы строки:

$$C_{2n}^{n\pm k} = C_{2n}^{n} \cdot \frac{n}{n+1} \cdot \frac{n-1}{n+2} \cdot .... \cdot \frac{n+1-k}{n \quad +k} \qquad (10)$$

или, в более компактной форме,

$$C_{2n}^{n\pm k} = C_{2n}^{n} \prod_{i=1}^{k} \frac{n+1-i}{n \quad +i} \ . \qquad (11)$$

Выражение (8) позволяет предложить следующую рекуррентную формулу для вычисления факториалов:

$$(2n)! = \frac{4^n (n!)^2}{\sqrt{\pi n}} \cdot \frac{n}{n+0,125} \cdot \frac{n^2}{n^2+0,0075} \qquad (12)$$

или

$$(2n)! = \frac{4^n (n!)^2}{\sqrt{\pi n}} \cdot \left(\frac{1}{1+\frac{0,125}{n}+\frac{0,0075}{n^2}+\frac{0,0009375}{n^3}} \right) . \qquad (13)$$

Следуя [2], покажем, что с помощью биномиальных коэффициентов можно вычислить триномиальные коэффициенты, а именно: триномиальные коэффициенты могут быть выражены через произведение биномиальных коэффициентов.

$$C_n^m C_m^k = \frac{n!}{m!(n-m)!} \cdot \frac{m!}{k!(m-k)!} = \frac{n!}{k!(m-k)!(n-m)!} \qquad (14)$$

Переобозначив переменные, получим:

$$C_n^m C_m^k = \frac{(a+b+c)!}{a!b!c!} , \qquad (15)$$

где $n= a+b+c;\ k=a;\ b=m-k;\ x=n-m.$

Таким образом, произведение биномиальных коэффициентов $C_n^m C_m^k$ есть не что иное, как триномиальный коэффициент, который появляется в триномиальной теореме:

$$(x+y+z)^n = \sum_{\substack{0\le a,b,c\le n \\ a+b+c=n}} \frac{(a+b+c)!}{a!b!c!} x^a y^b z^c = \sum_{\substack{0\le a,b,c\le n \\ a+b+c=n}} C_n^{b+a} C_{\mathrm{b+a}}^{a} x^a y^b z^c \qquad (16)$$

Учитывая следующее равенство для произведений биномиальных коэффициентов

$$C_n^m C_m^k = \frac{n!}{k!(m-k)!(n-m)!} = \frac{n!}{k!(n-k)!} \cdot \frac{(n-k)!}{(m-k)!(n-m)!} = C_n^k C_{n-k}^{m-k} \ , \qquad (17)$$

триномиальные коэффициенты можно представить в другой форме, несколько более удобной для суммирования:

$$(x+y+z)^n = \sum_{\substack{0\le a,b,c\le n \\ a+b+c=n}} \frac{(a+b+c)!}{a!b!c!} x^a y^b z^c = \sum_{\substack{0\le a,b,c\le n \\ a+b+c=n}} C_n^a C_{b+c}^{b} x^a y^b z^c \ . \qquad (18)$$

Обобщением биномиальных и триномиальных коэффициентов служат мультиномиальные коэффициенты, которые также могут быть представлены в виде произведения соответствующего количества биномиальных коэффициентов:

$$\frac{(a_1+a_2+...+a_m)!}{a_1!a_2!...a_m!} = C_{a_1+a_2+...+a_m}^{a_2+a_3+...+a_m} \cdot ... \cdot C_{a_{m-1}+a_m}^{a_m} \qquad (19)$$

Таким образом, имея достаточно точное выражение для биномиальных коэффициентов можно вычислять и триномиальные коэффициенты и, в общем случае, мультиномиальные коэффициенты.

Используя предлагаемое нами более точное приближение для биномиальных коэффициентов можно значительно повысить точность получаемых результатов, там, где раньше вынуждены были ограничиваться весьма приближёнными, как при использовании формулы Стирлинга, которая, как мы показали, даёт достаточно грубое приближение.

В работе [1] мы отмечали то, что числа в строках треугольника Паскаля удовлетворяют, по-видимому, бесконечному числу тождеств и свойств. В этой связи, обращает на себя внимание замечательное свойство произведений центральных элементов треугольника Паскаля [2]:

$$\sum_k C_{2k}^k \cdot C_{2(n-k)}^{n-k} = 4^n \ , \qquad (n \geq 0)\,.$$

Выводы и перспективы дальнейших исследований

В работе найдены значения параметров нормального распределения, аппроксимирующего распределение чисел в n-ой строке треугольника Паскаля. Вычислены значения нормированных моментов чётных порядков (с 4-го по 8-ой) и показано асимптотическое приближение (при возрастании номера строки) их величин к значениям, соответствующим нормальному распределению. Это свидетельствует о «нормализации» паскалевского распределения при больших значениях номера строки.

Показан волнообразный характер отклонений плотности паскалевского распределения от гауссовского в окрестности центральных элементов и «сглаживание» этого процесса для строк с большими номерами, что также свидетельствует о нормализации распределения.

На основе гауссовской аппроксимации получены высокоточные приближения для вычисления значений центральных элементов чётных строк треугольника Паскаля. Достаточно точные значения биномиальных коэффициентов позволяют вычислять триномиальные коэффициенты (с помощью произведения пар биномиальных коэффициентов) и, в общем случае, мультиномиальные коэффициенты (как произведения соответствующего числа биномиальных коэффициентов).

Перспективой дальнейших исследований может стать рассмотрение фрактальных свойств треугольника Паскаля.

Список использованной литературы

1. Абрамов Г.С. Нормальное приближение для распределения чисел в строках треугольника Паскаля / Г.С. Абрамов, И.М. Абрамов // Вестник ХНТУ. – Херсон: ХНТУ, 2014. – Вып. 3 (50), – С. 185-191.
2. Грэхем Р.О. Конкретная математика. Основание информатики: Пер. с англ./Р. Грэхем, Д. Кнут, О. Паташник. – М.: Мир, 1998.-703 с.
3. Феллер В. Введение в теорию вероятностей и её приложения: Пер. с англ./ В. Феллер. – М.: Мир, 1984. – 528 с.
4. Бондаренко Б.А. Обобщённые треугольники и пирамиды Паскаля, их фракталы, графы и приложения / Б.А. Бондаренко. – Ташкент, 1990. – 192 с.
5. Кузьмин О. В. Некоторые комбинаторные числа в обобщённой пирамиде Паскаля/ О.В. Кузьмин //Асимптотические и перечислительные задачи комбинаторного анализа. – Иркутск: Издательство Иркутского университета, 1998. – С. 90-100.
6. Кузьмин О. В. Обобщённые пирамиды Паскаля и их приложения / О. В. Кузьмин. – Новосибирск, 2000. – 64 с.